DOCUMENT RESUME

ED 059 727 LI 003 466

TITLE Use of Descriptors and Thesauri.
INSTITUTION United Nations Economic Commission for Latin America, New York, N.Y. Latin American Centre for Economic and Social Documentation (CLADES).
SPONS AGENCY United Nations Economic and Social Council, New York, N.Y.
REPORT NO CLADES-WG.1-L.2
PUB DATE 12 Aug 71
NOTE 60p.; (45 References); Paper prepared for Meeting on Modern Documentation Techniques, Santiago, Chile, 27 to 30 September 1971

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Classification; *Computer Programs; Electronic Data Processing; Foreign Countries; *Information Retrieval; *Programing Languages; Spanish Speaking; *Subject Index Terms; *Thesauri
IDENTIFIERS *Latin America

ABSTRACT

An attempt is made to provide a brief outline of the present situation with regard to the language used for computer retrieval. This language is increasingly based on descriptors and thesauri, which contain the key words for manual, mechanical or electronic data retrieval. After a review of the general problem, the document deals with techniques of preparing standardized lists of descriptors, considering them separately from thesauri which represent a higher and more complex level of elaboration which is not always required. Consideration is then given to the classification and construction of descriptors and thesauri, and it is also recommended that this vital aspect of data retrieval should be standardized. Lastly, an outline of some systems of descriptors and thesauri used in the international field is presented, ending with an explanation of the project which, as a subsystem, is being implemented by CLADES. This document should be considered not only as an information document, but also as a working document especially designed to help the participants in the Meeting on Modern Documentation Techniques to decide what standardization principles should be applicable to descriptors and thesauri in the Spanish-speaking region. (Author/NH)

ED 059727

UNITED NATIONS

ECONOMIC AND SOCIAL COUNCIL

LIMITED

CLADES/WG.1/L.2 ;
For participants only

12 August 1971

ENGLISH
ORIGINAL: SPANISH

ECONOMIC COMMISSION FOR LATIN AMERICA

MEETING ON MODERN DOCUMENTATION TECHNIQUES

Santiago, Chile, 27 to 30 September 1971

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

USE OF DESCRIPTORS AND THESAURI*

* Study prepared by the Latin American Centre for Economic and Social Documentation (CLADES), of ECLA for presentation at the Meeting on Modern Documentation Techniques.

LI 003 466

71-7-1937

TABLE OF CONTENTS

	Paragraphs	Page
ABBREVIATIONS		v
INTRODUCTION		vii
A. THE INTERRELATIONSHIP OF KNOWLEDGE	1- 8	1
1. Individual sciences and interdisciplinary sciences	1- 3	1
2. Disciplines and missions	4- 8	1
B. THE PROBLEM OF TERMINOLOGY STANDARDIZATION	9- 19	3
C. LISTS OF DESCRIPTORS	20- 78	5
1. Definitions	21- 27	6
2. Selection of descriptors	28- 29	6
3. Conceptual framework of selection	30- 32	7
4. Methodology of selection	33	8
5. Form of descriptors	34- 39	9
6. Methods of entering descriptors in the lists	40- 44	10
7. Methods of avoiding ambiguity	45- 48	11
8. Presentation of information	49- 50	12
9. The synopsis	51- 61	12
10. Data collection and retrieval	62- 67	14
11. Preparation of computer programmes	68- 70	17
12. Systems used for electronic data retrieval	71	18
13. Modification of lists	72- 78	18
D. THESAURI	79-103	20
1. Definitions	79- 83	20
2. General remarks	84- 89	21
3. Methodology and order	90- 91	22
4. Relations between descriptors	92- 93	23
5. Presentation of thesaurus	94- 99	25
6. General and specialized thesauri	100-103	26
E. CLASSIFICATION	104-154	28
1. The classic thesaurus	105	28
2., Subject headings	106	28
3. Enciclopaedic classifications	107-108	28
4. Universal Decimal Classification	109-110	29
5. Classification by categories or facets	111-118	30
6. Faceted thesauri	119-133	31
6.1 The Leyden thesaurus	121-122	32
6.2 The Thesaurofacet	123-129	33
6.3 The English classification of business studies	130-131	34
6.4 The ISIS thesaurus	132-133	35

/7. International

		Paragraphs	Page
7.	International standardization of classification	134-140	35
8.	The OECD Macrothesaurus	141-154	37
F.	UNITED NATIONS INFORMATION SYSTEMS	155-184	42
1.	United Nations Headquarters	158-162	42
1.1	The UNDEX system	158-159	42
1.2	Terminology lists	160	43
1.3	The Project Reports Information Subsystem (PRIS)	161-162	43
2.	The ILO Integrated Scientific Information System (ISIS)	163-166	43
3.	FAO Documentation Centre	167	44
4.	UNESCO system	168-170	45
5.	World Health Organization	171-172	45
6.	International Atomic Energy Agency	173	45
7.	International Telecommunication Union	174	46
8.	World Meteorological Organization	175	46
9.	United Nations Industrial Development Organization (UNIDO)	176-177	46
10.	International Bank for Reconstruction and Development (IBRD)	178-180	47
11.	Inter-American Development Bank project	181-183	48
12.	The UNCTAD system	184	48
G.	THE CLADES SYSTEM OF DESCRIPTORS	185-215	48
1.	Objectives	188-192	49
1.1	Efficiency	188	49
1.2	Simplicity	189	49
1.3	Low cost	190	49
1.4	Compatibility	191	49
1.5	Standardization	192	49
2.	Development model	193-202	50
2.1	Translation of Aligned List of Descriptors	193	50
2.2	User profiles	194-196	50
2.3	United Nations material	197-200	50
2.4	Government material	201	52
2.5	Other material	202	52
3.	Organization	203-215	52
3.1	List of descriptors	204-205	52
3.2	International co-operation	206	53
3.3	Disciplinary modules	207	53
3.4	Preparation of the thesaurus	208	53
3.5	Computer programmes	209-210	53
3.6	Codes	211-215	54
BIBLIOGRAPHY			1

ABBREVIATIONS

AID	Agency for International Development
CCAQ	Consultative Committee on Administrative Questions
CIDIA	Centro Interamericano de Documentación e Información Agrícola, Turrialba
CLADES	Latin American Centre for Economic and Social Documentation (United Nations, ECLA)
COSATI	Committee on Scientific and Technical Information (United States)
DIRS	Documentation and Information Retrieval System (FAO)
ECE	Economic Commission for Europe
ECLA	Economic Commission for Latin America
EURATOM	European Atomic Energy Community
FAO	Food and Agriculture Organization of the United Nations
FID	International Federation for Documentation
IAEA	International Atomic Energy Agency
IBRD	International Bank for Reconstruction and Development
IDB	Interamerican Development Bank
IICA	Instituto Interamericano de Ciencias Agrícolas. Turrialba
ILO	International Labour Organisation
ILPES	Latin American Institute for Economic and Social Planning
INIS	International Nuclear Information System (IAEA)
IOB	Inter-Organization Board for Information Systems and Related Activities (United Nations)
ISIS	Integrated Scientific Information System (ILO)
ISO	International Standards Organization
ITU	International Telecommunication Union
MEDLARS	Medical Literature Analysis and Retrieval System (United States/WHO)
NASA	National Aeronautics Space Administration (United States)
OAS	Organization of American States
OECD	Organization for Economic Co-operation and Development
PIRS	Programme Information Retrieval System (WHO)

/PRIS

PRIS	Project Reports Information Subsystem (United Nations)
STAR	Scientific and Technical Aerospace Reports (NASA, United States)
UNCTAD	United Nations Conference on Trade and Development
UNDEX	United Nations Documents Index
UNDP	United Nations Development Programme
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNIDO	United Nations Industrial Development Agency
UNISIST	World Science Information System (UNESCO and the International Council of Scientific Unions)
WHO	World Health Organization
WHOBRIS	WHO Biomedical Research Information Service
WMO	World Meteorological Organization

/INTRODUCTION

INTRODUCTION

This document is an attempt to provide a brief outline of the present situation with regard to the language used for computer retrieval. This language is increasingly based on descriptors and thesauri, which contain the key words for manual, mechanical or electronic data retrieval.

This is a difficult problem since, on the one hand, the languages must consist of vocabularies of many sciences and disciplines and, on the other hand, the systems must be placed at the service of new interdisciplinary sciences and of missions which depend on both the classical and the new sciences. Closely interrelated with this is the problem of terminology standardization, which is so important all over the world and particularly in the Spanish-speaking region, where the differences between countries have in many cases resulted in national or regional vocabularies that cannot easily be co-ordinated.

After a review of this general problem, the document deals with techniques of preparing standardized lists of descriptors, considering them separately from thesauri which represent a higher and more complex level of elaboration which is not always required.

Consideration is then given to the classification and construction of descriptors and thesauri, and it is also recommended that this vital aspect of data retrieval should be standardized.

Lastly, an outline of some systems of descriptors and thesauri used in the international field is presented, ending with an explanation of the project which as a subsystem is being implemented by CLADES.

The present document should be considered not only as an information document, but also as a working document especially designed to help the participants in the Meeting on Modern Documentation Techniques to decide what standardization principles should be applicable to descriptors and thesauri in the Spanish-speaking region.

A. THE INTERRELATIONSHIP OF KNOWLEDGE

1. Individual sciences and interdisciplinary sciences

1. Scientific knowledge, by the middle of the nineteenth century, was divided into disciplines - physics, chemistry, biology, sociology, etc. - which covered well-established conceptual areas and related to clearly defined natural systems, such as the atom, the molecule, living organisms or the various human activities. There were few points of contact between the disciplines; each scientific system operated in fairly strict isolation from other systems and had its own communication and information system.

2. With the appearance of interdisciplinary sciences, such as cybernetics and physiochemistry, and of a wide range of socio-economic activities, other intermediate disciplines began to be discovered and led to synthesis sciences. The world of specialization broke down in the nineteenth century and, while its validity is not denied, was reorganized very flexibly in line with generalized integration-oriented ideas.

3. Thus, information generated for a specific purpose in one science is often used for quite different purposes in other disciplines. The documentalist is therefore faced with a difficult problem of classification: he sometimes does not know according to what fundamental scheme he ought to label a segment of knowledge so as to make it available to all who need it, yet avoid redundant identifying, abstracting and announcing.

2. Disciplines and missions

4. The increasing importance of interdisciplinary fields has gone hand in hand with the extraordinary growth of mission-oriented science. For example, the exploration of space falls into almost all the traditional disciplinary fields. Space engineers or scientists who have carried out missions for the National Aeronautics and Space Administration (NASA) are also rocket chemists, guidance physicists, weightlessness physiologists and other specialists in widely different fields. The information they need should be collected in a mission-oriented information system, since what the

/physiologist learns

physiologist learns about weightlessness affects the engineer's desing of a space capsule, and is also useful to the basic physiologist interested in the kinesthetic sensory system. As indicated in the Weinberg report, often "the information originating in the discipline-oriented system must also find its way into the mission-oriented system".[1/]

5. This mission-discipline duality is also evident in the economic and social field. Balanced or integrated development, pollution control, ocean exploration and farming, or assitance in mitigating the effects of natural disasters, and, in general, the activities of international agencies entail complex missions involving knowledge of a wide range of disciplines integrated in many different ways.

6. The solution is not easy, since information must be collected by specialties or disciplines, i.e., conceptual and semantic units, while it must be integrated and processed with the maximum flexibility by missions of growing complexity.

7. Perhaps the best solution would be to create conceptual information modules based on the disciplines represented in semantic fields; they could be used as standardized and consolidated units to make up the complex systems of the missions.

8. In this area, general systems research [2/] could be usefully applied to the social sciences.

1/ See, under the head of discipline-mission duality, Alven Weinberg, "Science, government and information", The Growth of Knowledge: Readings on Organization and Retrieval of Information, ed. Manfred Kochen, New York, John Wiley and Sons, 1967, pages 42-44.

2/ See Ludwig von Bertalauffy, General System Theory (New York, G. Braziller, 1968); George J. Klir, An Approach to General Systems Theory (New York, Van Nostrand Reinhold, 1969); and the series entitled General Systems Yearbook, published by the Society for General Systems Research.

/B. THE

B. THE PROBLEM OF TERMINOLOGY STANDARDIZATION

9. The recent sweeping changes in information storage and retrieval, and particularly in the creation of new methodological tools, such as lists of descriptors and thesauri, adapted to mechanical and electronic information processing, have raised serious problems in regard to terminology differences and incompatibility between systems, not only in the sphere of national organizations, both public and private, but also at the level of international agencies. The lack of guidelines for the selection and use of vocabularies has given rise to a multitude of isolated information systems; some attempts have been made to remedy this situation in recent years through the publication and general acceptance of commonly observed principles for the establishment of thesauri.

10. In the United States, for example, the Committee on Scientific and Technical Information (COSATI) of the Federal Council for Science and Technology prepared guidelines 3/ summarizing the experience of important United States agencies, which have served as a tool for standardizing the structure of thesauri in that country.

11. Recently, UNESCO published more comprehensive guidelines for the establishment and development of monolingual scientific and technical thesauri for information retrieval, which were reviewed and discussed by persons of recognized competence in the field and were also examined at the International Conference on the General Principles of Thesaurus Building, which took place at Warsaw in March 1970. 4/

12. Moreover, the International Organization for Standardization (ISO) has set up a Technical Committee, known as ISO/TC 37, Terminology (principles and co-ordination), with the mission of finding out and formulating general principles on terminology and terminological lexicography. The recommendations of this Technical Committee deal with vocabularies of terminology, procedures

---

3/ COSATI, Guidelines for the Development of Information Retrieval Thesauri (first edition), Washington, D.C., September 1967.

4/ UNESCO, Guidelines for the Establishment and Development of Monolingual Scientific and Technical Thesauri for Information Retrieval (document SC/MD/20), Paris, 6 July 1970.

/for preparing

for preparing national or international standardized vocabularies, principles and criteria of value for the standardization of concepts, terms and their definitions, and layout of monolingual and multilingual vocabularies, including lexicographical symbols.[5/]

13. Concurrently, UNESCO and the International Council of Scientific Unions (ICSU) are jointly exploring the possibility of establishing a world science information system (UNISIST), and the feasibility study initiated in January 1967 has recently been published.[6/]

14. Recommendation 4 of group I on tools of systems intercommunication deals with subject specification in scientific information systems which involve "not only epistemological and practical problems relating to the classification of science during a period of vigorous growth and dramatic change, but also problems relating to the standardization of scientific terminology as among countries and as among the multiple fields of science for the applied purpose of developing controlled vocabularies or thesauri for the information systems, and of progressing towards the automatic indexing of documents".[7/]

15. Clearly, the feasibility of any world information system - scientific, scientific-technological or socio-economic - depends primarily on whether or not its components are compatible; it is equally clear that its components cannot be compatible unless common standards are adopted, which are widely accepted and applied, and which should include, first and foremost, those relating to the compilation and use of lists of descriptors and thesauri.

5/ The ISO recommendations of terminological interest include: Vocabulary of terminology (ISO/R 1087), Guide for the preparation of classified vocabularies (Example of method) (ISO/R 919), Naming principles (ISO/R 704), International unification of concepts and terms (ISO/R 860), Layout of multilingual classified vocabularies (ISO/R 1149), Layout of monolingual classified vocabularies (in course of preparation), Lexicographical symbols (draft ISO recommendation 1951), and Symbols for languages, countries and authorities (ISO/R 639).

6/ UNISIST, Report on the feasibility study on a world science information system by the United Nations Educational, Scientific and Cultural Organization, and the International Council of Scientific Unions (Paris, UNESCO, 1971); and UNISIST, Synopsis of the feasibility study on a world science information system, by the United Nations Educational, Scientific and Cultural Organization, and the International Council of Scientific Unions (Paris, UNESCO, 1971).

7/ Ibid., p. 45

/16. In

16. In Latin America, and in Spanish- and Portuguese-speaking countries in general, such compatibility is favoured by the existence of the two broad language areas and their relative backwardness in adopting modern indexing and bibliographical data retrieval techniques. The lack of structure is, at the moment, a certain advantage which can be fully exploited by establishing common principles and technologies that must greatly facilitate the regional exchange of information and relations with other language areas.

17. In this document on descriptors, thesauri and their use, the UNESCO standards have been adapted to the Spanish, account being taken of the work performed on the subject by distinguished Latin American research specialists; since it represents a synthesis, reorganization and adaptation of the available material, however, no indication is given of the source of the information in each item.

18. It is hoped that the perusal of these norms and their practical appraisal by Spanish-speaking research specialists will lead to constructive criticism as a basis for improving them in the future and converting them into a widely applied tool in this language area.

19. With a view to achieving regional co-ordination in this field, the best course would be to send information to CLADES (ECLA, Casilla 179-D, Santiago, Chile) about any plan to establish lists of descriptors or thesauri or any studies published on these subjects, and about any experience that is acquired, for dissemination and subsequent collective appraisal.

C. LISTS OF DESCRIPTORS

20. The creation and verification of a list of descriptors constitute the first stage in the preparation of a thesaurus. From a practical point of view, it is advisable to test the lists of descriptors for sufficient time to ensure their validity in relation to the proposed objectives. Frequently, when there is a limited amount of documentation, it will be sufficient to establish an adequate list of descriptors without it being necessary to undertake the much more arduous task of preparing a thesaurus.

/1. Definitions

1. Definitions

21. The descriptors may be defined as terms formed by one or more key words which summarize or denote a concept, with which a systematized vocabulary is built up that is used for indexing or retrieval of documents or information in general, within a given system of information.

22. Descriptors are supplemented by identifiers or indicators which give more depth to the index. Identifiers are terms which define temporal, spatial or modal circumstances, such as periods of time, geographical locations, names of institutions, lists of products, etc. They should be put on separate lists and codified when necessary.

23. The United Kingdom Classification Group defines the list of descriptors as a controlled terminological list that serves as a model for their use in indexing.[8/]

24. Descriptors may be simple and compound. Simple descriptors are made up of a single world, and compound descriptors of two or more words which always express a unitary concept. An attempt must be made to synthetize the concepts in the least possible number of words and to avoid prepositions and articles.

25. When the descriptors are prepared for computerized retrieval, the maximum number of characters they should possess - including blank spaces - should be determined in advance.

26. Candidate descriptors are those words which are selected initially to evaluate their use as descriptors.

27. Lists of descriptors may be generalized or common, when they cover the principal domains of learning, or specialized, if they refer to a given domain or areas marginal to that domain.

2. Selection of descriptors

28. To help choose adequate descriptors, an evaluation must be made of the frequency of use and the degree of understanding of the descriptors by the user, giving preference to concrete over abstract terms, and avoiding esoteric or outmoded terms.[9/]

29. There is some discussion as to whether the value of the descriptors should or should not be in direct relation to their repetition in the literature, since

---

8/ Definition approved by the United Kingdom Classification Group at its meeting No. 173 of 20 July 1970.(International Federation for Documentation (FID), News Bulletin, vol. 20, No. 9, 15 September 1970, The Hague.)

9/ J.W. Metcalfe, Subject classifying and indexing of libraries and literature, New York, The Scarecrow Press, 1959. /some authors

some authors consider that the words which best define a special field are those that appear most frequently, excluding generic terms and non-significant terms, while others consider that the value of the meaning of a word fluctuates in inverse proportion to the frequency of its use. Both these positions are too extreme. Although statistical analysis is very important to determine the retrieval value of a term, it is also true that some terms that are not in common use have a very high value for the user, and their retrieval is necessary.

3. Conceptual framework of selection [10/]

30. The criteria of selection will vary according to:

(a) the proposed structure of the list or thesaurus (systematic, conceptual or alphabetical listing, graphic display of relationships, facets, etc.);

(b) the purpose for which they are used (manual, mechanical or electronic compilation of information, cross-indexes, etc.);

(c) the background to the project (gradual build-up to the completion of electronic processing; introduction of a new domain, for instance, interdisciplinary areas for which there were no previous classification schemes; existence of well defined groups of users and subject specialists; extensive literature, etc.).

31. The selection of descriptors must be carried out on an experimental and not a theoretical basis, that is, by considering what terms are being used concretely in the programmes or work being carried out, the fundamental terms that appear repeatedly in the literature, classification schemes, lists of interdisciplinary terms, glossaries of the field in question, etc.

32. Relative frequency of use and exactitude of meaning are two complementary elements for the selection of descriptors. When both criteria are combined, the most useful descriptors are obtained.

10/ As has been mentioned, these guidelines for the construction of lists of descriptors and thesauri follow as closely as possible the guidelines given in UNESCO, Guidelines... op.cit. pages 5 et seq.

/4. Methodology

4. Methodology of selection

33. There are four distinct steps in the selection of candidate descriptors: collection, verification, evaluation and final choice.

(a) Collection

Potential users and subject specialists should first be consulted, the former in order to define the scope of users' interests and the latter to define the scope of the literature that is to be analysed (work programmes, technical works, specialized magazines, existing classification schemes, and any other source related to the proposed objective). The collection should be set out provisionally in alphabetical lists and semantic fields. The first facilitate the search for and verification of the existence or non-existence of candidate descriptors, while thought associations that result from their organization into groups give rise to many candidate descriptors that may be very valuable for information retrieval.

(b) Verification

Once the preliminary collection has been completed, the authenticity of the selected cantidate descriptors should be verified. To that end, there must be verification of the opinion of subject specialists, current usage in the literature, and the relevant dictionaries, indexes and standardized vocabularies.

(c) Evaluation

In evaluating the utility of candidate descriptors, reference should be made to their frequency as encountered in the bibliography and the literature on which the list is based; effectiveness in connoting and denoting the particular concept; anticipated frequency in requests for retrieval; relationship to descriptors already accepted; and appropriateness and authenticity as current terminology in the discipline concerned. None of these factors should be considered independently and particular attention should be paid to areas of peripheral interest, although in those areas the preciseness and specificity required of the descriptors are not the same as for the core subject.

/(d) Final

(d) Final selection

In all cases, descriptors should be selected in the last analysis, on the basis of their estimated effectiveness for retrieval purposes and their significance in the material to be indexed.

5. Form of descriptors

34. Word form. Once it has been decided to include a given term in the list of thesaurus, the form or the word must be made to reflect adequately the exact meaning intended.

(a) Spelling. The most widely accepted spelling of the word should be used. Where more than one spelling of a word is accepted (e.g. Spanish psicología and sicología) both forms should be included in the thesaurus, using cross-references. Alternatively, the Diccionario de la Real Academia Española or well-established technical dictionaries can be chosen to act as arbitrator.

(b) Translation. When foreign language terms and their translation coexist, they should all be included in the thesaurus and cross-referenced preferentially. (Example, Spanish computador, computadora, ordenador.)

(c) Transliteration. When a reference language is written in a different alphabet, the transliteration standards recommended by the International Organization for Standardization should be used. It is preferable to use a transliteration system which does not employ diacritical marks.

35. Noun form. The descriptor should be a substantive or a substantive form, modified by adjectives when necessary.

36. Number. In general, the plural form should be used for Spanish descriptors, when generic terms are involved. The singular form is used for specific material or property terms, process terms, proper names and disciplinary areas. When the singular and plural forms denote different concepts, both should be entered.

37. Abbreviations and acronyms

Abbreviated word forms and acronyms should be used only when their meaning is internationally established. Both abbreviated and unabbreviated forms should be displayed and cross-referenced. Sometimes the necessity for limiting the length of the descriptor entails the use of less well established abbreviations, in which case a scope note should be appended.

/38. Character

38. Character set. Since the majority of thesauri now being established will be used in connexion with electronic computers, it is advisable to use only the upper case format. Diacritical marks should be avoided for the same reason. The need for these restrictions will probably disappear in the near future, as progress is made in computer technology. The use of a computer may entail the limiting of the number of characters that a descriptor may have. In FAO, for instance, the number is reduced to twenty-eight, and in OECD to thirty-five.

39. Special characters and numerals. The only special characters allowed in descriptors are parantheses and unavoidable hyphens. Fullstops may sometimes be used. Any other non-alphanumeric symbols should be confined to scope notes, always within the limits of machine character availability. If the descriptors contain numeric elements, arabic numerals should be used, and their position should follow normal usage.

6. Methods of entering descriptors in the lists

40. Syntax. Compound expressions consisting of two or more words should not be inverted, e.g.: "political economy", not "economy, political". If inverted expressions are used, there should be a cross-reference from the forbidden term to the preferred term.

41. Punctuation. Punctuation marks do not usually appear in the descriptors, except parentheses, with the meaning indicated above. Where punctuation marks are omitted, however, it is advisable to include them in the scope notes. Hyphens can be easily omitted in Spanish, e.g., "económicosocial" instead of "económico-social".

42. Specialized and specific terms

(a) Specialized vocabularies

Whenever an international nomenclature or well-established standardized technical vocabularies exist, they should be used as descriptors or identifiers.

(b) Specific names

It is recommended that the names of unrelated specific entities be avoided; they could be included in a supplementary list of identifiers.

/(c) Specific

(c) Specific items

Descriptors representing generic, functional or structural concepts can be co-ordinated to denote specific items, and related by cross-references.

43. Alphabetization. Alphabetization rules should be clearly drawn up before any kind of ordering is attempted. The method of alphabetization depends on the size and structure of the domains covered by the list, the availability of machine processing, the existing programmes, etc.

The following alphabetization methods may be followed:

(a) Letter by letter (ignoring all spaces between words);

(b) Word by word (considering each complete word in turn);

(c) Computer sort (depends on the computer configuration and programmes used).

44. Synonyms and quasi-synonyms. In general, no language has any true synonyms, but when one term must be searched every time that another is searched, they should be treated as synonyms. Descriptors that overlap significantly or represent different aspects of the same property may be considered quasi-synonyms. Antonyms should be similarly treated. When synonyms, quasi-synonyms or antonyms are included, cross-references should be used.

7. Methods of avoiding ambiguity

45. In compiling lists of descriptors difficulties are encountered with terms which have more than one accepted meaning or whose meaning in a given context is different to that commonly encountered. In such cases the required meaning may be brought out by the use of the following methods.

46. Compound expressions. The use of modifying expressions to make clear the exact meaning associated with a given term is sometimes necessary, and in these cases the descriptors will consist of two or more words. Examples: "public economy", "per capita output".

47. Qualifiers for homonyms. Homonyms may be distinguished by the use of qualifying expressions placed between parentheses immediately after the term. Parentheses should be used only for this purpose. Example: beams (electromagnetic), beams (structural). Parentheses should not be used for synonyms. This method and the previous method are, in principle, mutually exclusive and should not be used together in the same list.

/48. Scope

48. Scope notes

A scope note is a brief explanation which accompanies a given descriptor in the list or in the thesaurus to indicate the scope of the term, but does not form part of the descriptor. It indicates the way in which the descriptor should be used and need not be a dictionary definition. Scope notes are used mainly for the two following purposes:

(a) To give the descriptor a given value or meaning:

Example:

Youth

Note: "Person aged 14 to 25 years"

(b) To establish the usage of a descriptor when it has several meanings:

Example:

Documentation

Note: "The process of storing and retrieving information in all fields of learning".

Documentation

Note: "The volume of documents assembled or available".

Documentation

Note: "The title of a family of publications".

Scope notes should always be used in connexion with abbreviations and acronyms.

8. Presentation of information

49. As a rule, the information contained in the descriptors is placed after the bibliographic description of the document.

50. Isolated descriptors. Isolated descriptors are often placed inmediately after the bibliographic description.

9. The synopsis

51. A study in depth of each document leads to the preparation of a synopsis which reflects the content of the document and complements the bibliographical information. The synopsis may be either integrated or not integrated with the descriptors.

/52. A

52. A synopsis which is not integrated is merely an abstract of the content of the document analysed, which serves as a guide for the human search and excludes any documents which are not relevant. In this case, the descriptors are separate from the abstract.

53. A synopsis which is integrated also contains words that are not retrievable; its text reads smoothly, it is more comprehensible and its presentation is more elegant.

54. A synopsis which is not integrated with the descriptors entails the double work of drafting the abstract and obtaining the descriptors and takes up more space in the computer's memory, although it may be useful in giving a fuller explanation of the content of a document which is retrievable with only a few descriptors.

55. On the other hand, a synopsis which is integrated with the descriptors, means a bigger over-all economy, both in the analyst's efforts and the computer's time; this also means a saving in time for the user, since the synopsis he must read will generally be shorter.

56. CLADES is trying out the type of synopsis that is integrated with the descriptors, using upper-case letters only, since there are no computer printing machines in Chile capable of producing lists in upper- and lower-case letters. The descriptors appear between slashes so that they are distinguishable for the computer, although they could be eliminated by proper programming.

57. The three following examples of synopses include a sample of the synopsis which CLADES proposes to use.

58. The ILO ISIS system.[11/]

29068 1967 12783-
MALDONADO VA
PARIS. UNIVERSITE. INSTITUT DES HAUTES ETUDES DE L'AMERIQUE LATINE
LE MARCHE COMMUN LATINO-AMERICAIN - ASPECTS DE L'INTEGRATION ECONOMIQUE EN AMERIQUE LATINE.
PARIS, 1967, XIV, 293 P. TABLES. (ITS - TRAVAUX ET MEMOIRES, NO. 19).

11/ Computer listing of the ILO Service of Bibliographic Information.

//THESIS/ON

/THESIS/ON PROBLEMS OF AND /OBSTACLE/S TO /ECONOMIC INTEGRATION/ IN /LATIN AMERICA/ - INCLUDES A SECTION ON /EEC/ EXPERIENCE, AND COVERS /ECONOMIC DEVELOPMENT/ IN / LAFTA/ COUNTRIES, /TRADE/, /INDUSTRIAL STRUCTURE/S, THE /BALANCE OF PAYMENT/S, /FOREIGN INVESTMENT/, /ECONOMIC POLICY, /INDUSTRIAL POLICY/, /TRADE AGREEMENT/S, /NATIONAL PLANNING/, /MONETARY POLICY/, /FISCAL POLICY/, /CREDIT/ FACILITIES, ETC. /BIBLIOGRAPHY/ PP. 287 TO 289, /IMF/ AND /UN/ MENTIONED, AND /STATISTIC/S.

FREN

59. The NASA STAR system[12]

N66-10552*# General Dynamics/Astronautics, San Diego, Calif.

WATER MANAGEMENT SUBSYSTEM SPECIFICATION FOR SPACE FLIGHTS OF EXTENDED TIME PERIODS. LIFE SUPPORT SYSTEM FOR SPACE FLIGHTS OF EXTENDED TIME PERIODS

J.A. Steele 30 Nov. 1965 31 p refs Revised (Contract NAS1-2934)

(NASA-CR-67600: GD/A-64-26211, Rev.A)
CFSTI: HC $2.00/MF$0.50 CSCL 06K

The water management subsystem specifications are presented for the equipment required in an operational water management program as an integral part of the life support system for space flight. The requirements of the subsystem components are summarized and the operational procedures for normal and emergency situations are outlined. Both the prototype and the flight systems are discussed. E.E.B.

12/ The Scientific and Technical Aerospace Reports (STAR) system of the National Aeronautics and Space Administration summarizes the documentatic on the exploration and study of space and on aeronautics and astronautics published all over the world. It contains some 12,000 descriptors covering thirty-four subjects or fields, such as, aerodynamics, aircraft, auxiliary systems, bio-sciences, biotechnology, chemistry, communications, computers, etc. The subjects have been clasiffied in accordance not with a scientific criterion, but with the practical interests of research specialists and technical experts using the system. The STAR consist of bibliographies and two-monthly information synopses, with indexes, and a cumulative six-monthly and annual index. (See J. Lasso de la Vega, Manual de Documentación, Editorial Labor, Barcelona, 1969, page 210).

/60. The

60. The CLADES system (experimental)

First page of the document (in preparation).

DIRECTORIOS EXISTENTES EN LA BIBLIOTECA DE REFERENCIA DEL CLADES

(First semester 1971)

1968

ASSOCIATION INTERNATIONALE DES ETUDES ET RECHERCHES SUR L'INFORMATION.
REPERTOIRE EUROPEEN DES ORGANISMES DE RECHERCHE ET DES CHERCHEURS DANS LE DOMAINE DE L'INFORMATION. REPERTOIRE EUROPEEN DES OUVRAGES PUBLIES DEPUIS 1966 ET DES RECHERCHES EN COURS CONCERNANT L'INFORMATION.
LAUSANNE, 1968. 113 P.

/*DIRECTORIO/DE /INSTITUTOS/ DE / INVESTIGACION/ Y DE /*PERSONAL INVESTIGADOR/ SOBRE /TECNICA DE INFORMACION/ EN/EUROPA/ /BIBLIOGRAFIA/ SOBRE INFORMACION DESDE 1966-
FRAN.

1968

AVICENNE P.
BIBLIOGRAPHICAL SERVICES THROUGHOUT THE WORLD, 1960-1964.
PARIS, UNESCO, 1968. 228 P. (UNESCO BIBLIOGRAPHICAL HANDBOOKS, 11)
/DIRECTORIO/ DE SERVICIOS DE /BIBLIOGRAFIA/ DE 83 PAISES, CUBRE 1960-1964 /INDICE GENERAL/ INDICE GEOGRAFICO/*LISTA DE PERSONAS/ Y DE /*INSTITUCIONES/
ING

1970

CAMARA DE COMERCIO DE BOGOTA. DEPARTAMENTO DE COORDINACION Y DESARROLLO.
ORGANISMOS ECONOMICOS COLOMBIANOS Y SUS SIGLAS.
BOGOTA, 1970. 163 P.
/*DIRECTORIO DE /INSTITUCIONES/ DE /ECONOMIA/ EN /COLOMBIA/ INDICE GENERAL/*LISTA DE SIGLAS/ INDICE GEOGRAFICO/
ESP

61. Preparation of lists. Once the descriptors have been selected and elaborated, they should be ordered logically by semantic fields or facets and the general alphabetical index should be finally prepared. Lists of qualifiers and supplementary codes of acronyms and other elements conclude this important stage and convert the lists of descriptors into a valuable tool for indexing and information retrieval.

---

*/ Descriptors preceded by an asterisk (*) are used by CLADES. Those without an asterisk will be found in the Aligned Descriptor List of OECD, ed. 1969.

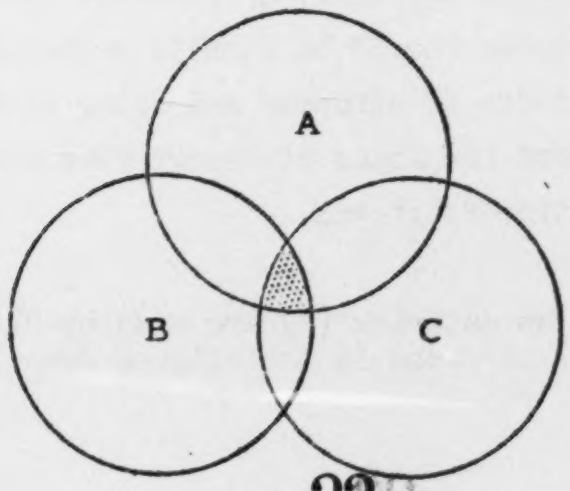
10. Data collection and retrieval

62. Descriptors, together with bibliographical description, constitute the units of data retrieval, and accordingly the recall ratio, or the relationship between the information required and that obtained within a given system, depends on the selection and structure of these units. This ratio is not normally equal to one since, however perfect the system, there will be documents that do not appear in the retrieval, although they are relevant to the search, while others that are inadequate will appear.

63. The great diversity of methods put forward for the organization of thesauri and lists is aimed at making that ratio as close as possible to one, to obtain as far as possible only those documents that the investigator requires.

If the descriptor or combination of descriptors utilized is of a very broad generic group, the reply to the search will give a greater number of parasitic documents (or "noise" in terms of the information theory), which do not answer the question. On the other hand, if the generic group is too restricted, documents of interest will be excluded. It will be difficult to achieve the equilibrium needed to obtain the required information and to reduce the number of documents containing parasitic information; this will depend to a large extent on the design and permanent checking of the system.

64. The descriptors make it possible to reduce the scope of the search by defining it in terms of the required exactitude through the Boolean combination of terms, as the following diagram shows:



/The letters

The letters denoting each of the three circles A, B and C represent descriptors, while the shaded area represents the information required. The other areas represent parasitic information in relation to the information required.

65. If it is wished, for instance, to search for the information that exists on the "impact of / technological change / on / economic development/ in / Mexico"/, on the basis of the descriptors A ("technological change"), B ("economic development") and C ("Mexico"), the system will search the whole bibliography that contains any of those descriptors - or any others that may be established - and will give as a result only those documents that contain all the descriptors in question.

66. The user must formulate his request for information in concrete terms, and he should express his interests clearly and describe as precisely as possible the characteristics of the problem, the bibliographical background and any other limiting factor. As a rule, this request may be presented in non-technical language.

67. The documentalist, in turn, will have to translate that request into descriptors and will have to present the logical formula or formulas to permit their automatic retrieval.

11. Preparation of computer programmes

68. If it has been decided initially to carry out electronic processing and retrieval, programming should be initiated at the same time as a start is made on collecting candidate descriptors. The preparation of a computer system is, of course, a complex and relatively expensive process, especially in the indexing phase, although it only requires the work of programmers with a certain amount of experience in the field, since a variety of modules and programmes are available to construct many systems of very diverse capacity.

69. Languages. Any general language, that is not of a strictly mathematical nature can be used for programming, such as the Cobol or Assembler languages, or a mixture of both, or such as the PL-1, which is being more widely used of late and has been adopted by the United Nations PRIS system. There are also other more specialized languages, such as LISP, the advantages and drawbacks of which cannot be discussed here.

/70. Programmes.

70. Programmes. CLADES has already begun preparing modules and basic programmes for elaboration into a practical system, of the greatest possible simplicity and with various levels of complexity, which it hopes to develop in a relatively short time and which it will make available to interested governments and official organizations, as a form of technical assistance. (See the chapter entitled "The CLADES' system of descriptors".)

12. Systems used for electronic data retrieval

71. Many models are used. Within the United Nations family alone, each organization uses a different configuration; these are described in the Henderson report, which covers nearly all the specialized agencies which carry out electronic processing of a considerable amount of data.[13/] At the present time, considerable efforts are being made to standardize United Nations systems for the electronic retrieval of data, to which end a standardizing agency, the Inter-Organization Board for Information Systems and Related Activities (IOB), has been set up in Geneva.

13. Modification of lists

72. As practice has shown, the lists of descriptors have to be revised periodically so that new elements that arise may be incorporated and outmoded terms may be eliminated.

(a) Periodical verification

The validity of the descriptors must be verified from time to time, both for indexing and retrieval purposes. When a search does not uncover the volume of information that is required or does not cover the volume that is considered to exist in the data base, a critical examination of the descriptors used in comparison with those that should have been used should be carried out.

(b) Elimination of descriptors

The elimination of descriptors may be either by default or due to excessive coverage.

13/ A.M. Henderson, Electronic data processing in the United Nations family of organizations, Ottawa, Auditor General of Canada, 1970, two volumes.

/(i) Elimination

(i) Elimination by default. Descriptors that are very seldom used must be eliminated or replaced by a more common term, provided that the non-usage is not merely due to the lack of documents dealing with the concept in question. In general, the use of a preferential term to indicate the substitution is more practical than complete elimination.

(ii) Elimination due to excessive coverage. If the same descriptor is used to describe many items, its specificity is lost because its application is too general, and the possibility of breaking down the concept into two or more descriptors should be studied.

The search strategy that is employed for retrieval determines to a certain extent the procedure that must be followed when a descriptor is used too little or too much. If it is the least specific descriptor that is being sought, it may ultimately be retained.

73. New descriptors. Indexers and users must continuously investigate the possibility of using new descriptors that represent new concepts or different facets of old concepts. It is recommended that indexers should first test the proposed descriptors for a certain length of time, before including them definitely in the list or thesaurus.

74. The date of inclusion must be mentioned, and indexers should not use them before that date.

75. When a new descriptor is added, its consequent interrelationships with the other descriptors must be identified and introduced where appropriate.

76. New descriptors must be included in the list or thesaurus by groups, either as additions or when a new edition is published, which does not exclude their use by indexers.

77. There must be a central authority to examine and decide as to the acceptability of all suggestions for the use of new descriptors.

78. It must be borne in mind that a thesaurus never reaches a final shape, since both its form and size are functions of time.

/D. THESAURI

D. THESAURI

1. Definitions

79. Definitions of thesauri all agree that they consist of a defined and controlled set or interrelated terms that are used for the retrieval of information within a given range.

80. A standard definition would describe the thesaurus as a controlled terminological list of conceptually associated terms for use in data retrieval systems in connexion with post-co-ordinated indexing.[14/] Similarly, the thesaurus has been defined as a structured lexicon designed for the selective indexing and retrieval of information.[15/]

81. UNESCO defines the thesaurus as "a controlled and dynamic vocabulary of semantically and generically related terms which comprehensively covers a specific domain of knowledge".[16/]

82. The thesaurus can also be defined in general terms as a dictionary of words of a natural language classified by semantic affinities in accordance with certain criteria, i.e., a dictionary of concepts. In the strict sense as applied to document automation, the thesaurus is a dictionary of absolute or conditional equivalences between the words or expressions of a natural language and the descriptors of a documental language.[17/]

83. Professor Silva describes the technology of descriptors as the set of operations - whether simple or complex, manual, mechanized or automatic - into which the modern process of providing and receiving information through descriptors can be broken down; he considers that the term "descriptor" refers above all to the automatic or electronic treatment of information, that the term "keyword" relates to the use of mechanical

---

14/ Definition of the United Kingdom Classification Research Group adopted at its 17th., meeting on 20 July 1970.

15/ Luis Roberto Barbosa de Oliveira, Thesauri - Sua Fundamentação (Rio de Janeiro, Instituto de Documentação, Fundação Getulio Vargas, 1970).

16/ UNESCO, Guidelines, op.cit., p.2.

17/ R.C. Cros, J.C. Gardin and F. Levy, L'automatisation des recherches documentaires: un modèle général, le Syntol, 2nd., revised and enlarged edition (Paris, Gauthier Villars, 1968).

/means and

means and that the techniques of co-ordinated term and subject headings correspond to the stage of technical evolution that is mainly manual, although he does recognize that descriptor thesauri are also subject to semi-mechanized or totally mechanized use, etc.[18/]

2. General remarks

84. As Sharp points out,[19/] a thesaurus can be used for one of two quite opposite purposes, either to broaden a request by allowing the looking up of terms related to the original request, or - more legitimately, in his view - to restrict the request by subsuming the terms suggested by the user under a given set of terms which are then exclusively used in the matching stage - a process which is carried out by the person responsible for the matching who looks up the terms assigned by the indexer or retriever in a pre-prepared thesaurus.

85. One of the greatest virtues of the thesaurus is its ability to assimilate immediately the neologisms and special terms that abound in the ever-growing field of basic and applied research. Every advantage should be taken of this quality, in combination with the use of explanatory notes and cross-references. Particular care should be taken with terms whose connotations have altered with time or whose sense varies from one country to another. When terms relating to various semantic fields are included, the appropriate cross-references are called for.

86. From a practical point of view, a thesaurus is not a set of strict norms but a guide for the processing and retrieval of information material, which means that the nature of the material processed and the interests of the users should have priority over any abstract or rigid norm.

18/ Benedicto Silva, Curso de Introdução a Tecnologia des Descritores: Origem e Evolução des Descritores (Rio de Janeiro, Instituto de Documentação, Fundação Getulio Vargas, 1970), pp. 14 and 30.

19/ J.R. Sharp, Some Fundamentals of Information Retrieval (1965), as quoted in J.P. Henley, Computer-Based Library and Information Systems (London, MacDonald, 1970), p. 33.

/87. Thus,

87. Thus, unlike classification systems, thesauri do not necessarily have to employ categories; they do, on the other hand, have to be as flexible as possible so as to satisfy the changing requirements of the users and the rapid development of a constantly evolving science and technology.

88. Consequently, a thesaurus should not be an objective of a documentation centre but merely a working instrument to be employed when the field or speciality is not yet properly covered by appropriate controlled vocabularies or when the volume or utilization or the information so demands.[20/]

89. It should not be forgotten that thesauri are expensive to build up from both the technical and the financial standpoint and that they demand much more attention, time and effort than might be thought at first sight.

3. Methodology and order

90. The structuring of a thesaurus' lexicon should be carried out in the following stages:

(a) Re-compilation of the technical terminology;

(b) Reduction of the universe to be structured to a minimum vocabulary by means of synonyms or any other form of semantic derivation;

(c) Creation of a compiling grammar for defining all the terms in the lexicon or for setting them out logically on the basis of the terms of the minimum vocabulary.[21/]

91. In the view of Professor García de Miranda, there are three kinds of order for terms or descriptors, in increasing degree of flexibility and complexity:

(a) The lineal order, in which the elements of documentation are set out in sequence in accordance with a criterion which is independent of their semantic situation - alphabetical order, number of words, or any other criterion;

20/ Zulma P. de V. Courrege, O Thesaurus (Rio de Janeiro, Instituto de Documentação, Fundação Getulio Vargas, 1970), p. 13.

21/ Luis Roberto Barboza de Oliveira, Thesauri, op.cit., p. 7.

/(b) The

(b) The two-dimensional ramified order, which consists of categories or sub-categories, classic examples of which are the classifications of Dewey and of Otlet and Lafontaine (also known as branched order);

(c) The multidimensional reticulated order, which takes into account both the hierarchy of the elements of a (ramified) repertory and the associations between these elements - this is above all the order employed in thesauri, while the lineal order is used in simple lists of descriptors and the ramified order for hierarchical classifications.

In the opinion of the same author, it is a specific characteristic of a thesaurus to be a dictionary of terms and ideas.[22/]

4. Relations between descriptors

92. The most important function of thesauri is to serve as a tool for information retrieval, bringing into evidence the interrelationships between individual descriptors, together with the synonyms that enable them to be arrived at on the basis of natural language. These ... interrelationships are demonstrated by simple references, among forbidden terms and descriptors, or through cross references, which always have the property of reciprocity, as the relationship between one descriptor and another, and vice versa, should in all cases be indexed.

93. These interrelationships are of three types: preferential, hierarchical and affinitive.

(a) Preferential relationship

This reference is used to refer from a forbidden term, that is, a non-descriptor, to a descriptor, and is used in the following cases:

(i) Real or attributed coincidence between two or more terms (synonyms or quasi-synonyms);

(ii) Selection of one descriptor as representative of a group of descriptors (which may include antonyms);

(iii) Where different spellings of the same word exist (for instance, Spanish sicología and psicología).

22/ Professor Antonio García de Miranda, Aspectos semanticos do thesaurus (Rio de Janeiro, Instituto de Documentação, Fundação Getulio Vargas, 1970).

/These references

These references have the code: use.

Examples: plane, use aeroplane; (Spanish) sicología, use psicología.

The inverse reference is: used for.

(b) Hierarchical relationship

Hierarchical relationships are used to exhibit relative degrees of specificity within a category of descriptors all of which belong to a particular generic group. This relationship is not based on the use or application of a concept or term, but on the position of the descriptor within a given class of concepts. Certain terms may be members of more than one hierarchical class. Where any hierarchy has more than two levels, the cross references for all levels should be completed for each descriptor. In general, all concepts which are sub-divisions of a broader concept should form part of a hierarchical chain.

It should be borne in mind that when the vocabulary is very varied and complex - as is the case with information on current events - 23/ and when the terms are not univocal - as is the case in the economic and social fields - it is not always possible to determine the hierarchical relationships. This does not normally happen in the more limited and precise fields of technology and the natural sciences, and so hierarchical relationships in economics and sociology should be indicated only when they are very obvious.

These relationships must always be expressed as cross-references, in the following forms:

Broader term (BT);

Narrower term (NT).

(c) Affinitive or collateral relationship

The affinitive relationship is employed to refer from one descriptor to others that are closely related to it in concept but are not consistently related either hierarchically or preferentially.

This relationship is expressed by the code: related term (RT). Example: education, RT learning, and learning, RT education.

23/ See the New York Times, Thesaurus of descriptors, second edition, vol. 1 (New York, the New York Times Co., 1969).

/It is

It is advisable to use this type of reference when there is some doubt as to the hierarchical level or when it is considered that both terms are on the same hierarchical level.

Another even simpler system of non-hierarchical cross references consists of the following: See and See also. They are used in the following way:

Reference See.

This is used in all preferential references from forbidden terms or non-descriptors, to descriptors. In other words, it leads from a term that is not used as an entry in the system to an equivalent that is used as an entry.

Reference See also.

This leads from a descriptor that is used as an entry in the system to one or more other descriptors which are also entered, under which information relating to the first descriptor is included.

In general, these references lead from one descriptor to another on the same hierarchical plane, referring to different aspects of the same theme or to different peripheral themes (ROADS, see also traffic). They may also lead from broader terms to narrower terms (HOUSING, see also real estate) or vice versa.

5. Presentation of thesaurus

94. A thesaurus should be more than just a list of descriptors, and should be composed of the following parts:

95. Systematical listing. The descriptors are grouped in general classes or categories, within which relationships - hierarchical, logical or of any other kind - are established between the terms.

96. Alphabetical listing. This is combined with the first and can be expanded by permuted alphabetical lists (such as KWIT, KWIC and KWOC). The alphabetical listing has the advantage that it can be consulted directly to discover the presence or absence of the term, and at the same time the introduction and correct positioning of new descriptors is easy.

/97. The

97. The combination of these two methods facilitates the search for a given term in the alphabetical listing, while the systematical listing makes it possible to establish associations of terms and to broaden the search in peripheral or related areas.

98. Graphic display, using co-ordinates or arrows. Here, too, the descriptors are organized in semantic groups and are interrelated by various means. For instance, each descriptor is placed within a system of co-ordinates and associative relationships are shown by bi-directional arrows, hierarchical relationships by uni-directional arrows pointed to the more specific descriptor. Preferential relationships may be indicated by brackets and arrows pointed to the term that should be used. These relationships may be represented two-dimensionally, assigning two dimensions to each aspect and combining them multi-dimensionally.

99. The thesaurus is supplemented by lists of synonyms or quasi-synonyms, and codes for countries, institutions, magazines, etc., which facilitate the work of manual, mechanical or electronic retrieval.

6. General and specialized thesauri

100. The proliferation of lists of descriptors and thesauri designed to meet an organization's temporary or particular needs could lead to a chaotic situation which might prevent or impede access to valuable sources of informations. The large number of complex classification systems, which sometimes reflect a great deal of phantasy and little practical sense or have not been sufficiently put to the test of experience, constitute perhaps the most serious danger for intercommunication of the specialized thesauri which are being prepared in many different places without reciprocal co-ordination or knowledge. It is

/therefore of

therefore of the utmost urgency to standardize the controlled vocabularies from the standpoint of both uniformity of terminology and the methods used in elaborating them.

101. It is increasingly necessary to have a general thesaurus or macro-thesaurus which will operate as the common rule for specialized thesauri and micro-thesauri and will reflect, in the form of subgroups, the most important and widely used concepts of the various economic, social and technological sciences and activities.

102. This is not an easy task, of course. The experience obtained through different versions of the OECD Common List of Descriptors (see E.8, The OECD macro-thesaurus) shows that the work must be performed on an international and interdisciplinary basis by highly-qualified experts, such as those who have been engaged in the task.

103. Little would be gained, however, if the specialized thesauri were not co-ordinated with the general thesaurus, in order to obtain an integrated system that will permit access to the different specialties from a central point, it being of the utmost importance that as many economic and social development organizations as possible take part in elaborating this instrument, so that all the semantic fields will be duly represented in their broadest expression.

/E. CLASSIFICATION

E. CLASSIFICATION

104. One of the most widely discussed questions is what system of classification should be used to organize retrieval and information tools, such as lists of descriptors and thesauri. Some of the systems proposed are examined briefly below.

1. The classic thesaurus

105. The original purpose of the thesaurus was to arrange terms by semantic or conceptual fields, as in Roget's classic thesaurus, in order to facilitate the search for synonyms and related concepts. The essential aim was to broaden the association between concepts, facilitate the search for peripheral terms around a central theme, and to pass from one conceptual field to other related areas. These conceptual associations may be purely logical - establishing a relationship between the components of a group, a part and the whole, the genus and the species, the aspects of a process, the components of a machine, etc. - or metaphorical and symbolic. Today, the thesaurus is a complex system used especially for electronic information retrieval.

2. Subject headings

106. The standardized lists of subject headings are used only for the general classification of documents or superficial indexing, as opposed to the more thorough analysis of content made in the lists of descriptors and specialized thesauri. It is considered that a standardized list of general subject headings may be useful when trying to locate and transfer large blocks of information, rather than specific documents or data.[24/]

3. Encyclopaedic classifications

107. The UNISIST Working Group on documentation research requirements examined the various encyclopaedic classifications which compete for world recognition as accepted rules for classification by subjects, e.g., Dewey's Decimal Classification, the Library of Congress List of Subject Headings,

---

24/ Study Report on the Feasibility Study on a World Science Information System, op. cit.

/the Universal

the Universal Decimal Classification, etc. After weighing their advantages and shortcomings, it concluded that no clear answer could be given to the more controversial question of the lack of total or local adaptation to the content and structure of the UDC divisions, and considered that additional studies and experiments were required to measure the potential value of UDC in its present condition of single world list of special subject headings for general classifications or superficial document indexing.

108. The experiments carried out to demonstrate that encyclopaedic classifications may also be used as description languages for thorough indexing and mechanized retrieval do not prove that they are desirable, since there are sound arguments for preserving the distinction between classifications by means of rigid boxing, on the one hand, and completely flexible descriptor languages, on the other.[25/]

4. Universal Decimal Classification

109. The use of the Universal Decimal Classification (UDC) in the arrangement of thesauri has ardent supporters, such as Lasso de la Vega, who considers that UDC has become an excellent system for the elaboration of any topic, subject or item,[26/] and equally ardent detractors, such as Barbosa de Oliveira, who believes that when searching for a specific subject, retrieval by means of UDC is almost like guesswork, if not actually impossible, both because the name given to the subjects searched must be exactly the same as those given to their subjects by the authors of UDC, and because the whole question of the setting or environment of each specific item is arbitrary and artificial, since the system constitutes a semi-aleatory group of sciences and technologies whose genealogical tree is divided into a succession of rigid categories each comprising ten units. He therefore believes that the possibilities of retrieving information by means of UDC will rapidly be disappearing as specialization in the different subjects increases.[27/]

25/ Ibid.

26/ J. Lasso de la Vega, Manual de documentación, op. cit., p. 197.

27/ Luis Roberto Barboza de Oliveira, Thesauri - Sua Fundamentação, Documentation Institute, Getulio Vargas Foundation, Rio de Janeiro, 1970, pp. 5-6.

/110. Clearly

110. Clearly, if a classification such as UDC, which is of very general application, were adopted on an international scale, there would be some advantages as regards compatibility of the systems, but it is true that at the present stage it is doubtful whether it would be wise to use a system which, in practice, is seen to lack the necessary flexibility and is of highly complex and specialized application.

5. Classification by categories or facets

111. This classification covers a wide range of systems from that which draws a sharp distinction between classification by categories or facets and thesauri, to an attempt to integrate the two systems. Likewise, proposals have ranged from a system of five single categories to considering any group of descriptors as a facet, whereby this concept takes on maximum relativity and flexibility.

112. The first colon or faceted classification - based on the Aristotelian concept of categories - which was tried out and elaborated by Ranganathan 28/ more than twenty years ago as applicable to libraries, but not to documentation in depth, covers the categories of personality (whose significance has not been defined), matter, energy, space and time.

113. This classification has been the object of various criticisms, including that of Eric de Grolier,29/ who considers it useless to compare the theory on which it is based with its practical application since, if the former is seen to be very ambitious and to have a philosophical and even highly dogmatic bias, the latter appears to be singularly empirical and at the same time extremely arbitrary.

28/ S.R.Ranganathan, Classification and Communication, Delhi, 1951; Philosophy of Library Classification, Copenhagen, 1951; and other studies.

29/ See Eric de Grolier, Théorie et pratique des classifications documentaires, Paris, 1956; Le progrés et l'avenir du langage classification, The Hague, 1958; and Etude sur les catégories générales applicables aux classifications documentaires, Paris, 1962.

/114. Eric

114. Eric de Grolier divides the categories into constants (time, space and action) and variables (substance, analytical organ, synthetic organ, property, formula and organization).

115. Vickery suggests an even more complex combination which he symbolizes as follows: substance, product, organism (P); part, organ, structure (O); component (C); property and measurement (Q); object of action, raw materials (R); action, operation, method, behaviour (E); agent, space and time (ST).

116. The Classification Research Group, established in London in 1952, has worked intensively on this system, improving some of its aspects.

117. The limited number of facets is useful where fairly rigid systems must be applied, as in technological classifications or the system of the Highway Research Board of the United States, which is concerned with a means of transport with highly stable elements.

118. However, the complexity, diversity and rigidity of these systems makes them unsuitable for use on an international or regional scale, since at this level standardization - and at the same time flexibility - are increasingly important requirements.

6. Faceted thesauri

119. Classification by facets differs from the thesaurus in one significant respect: while facet classification unites fundamental concepts of different groups so that, for example, an operation and its associated agents are generally shown together and share common elements of notation, in the thesaurus things are separated from acts, etc., so that the classifier is free to combine those different concepts in accordance with the formula he chooses.

120. There have, however, been several attempts to combine both these ideas of conceptual organization in order to make up a single structure by means of faceted thesauri, which relate a facet classification with a combined alphabetical index to a hierarchically organized thesaurus, so that it would amount to the same thing to consider the thesaurus as a classification index or the facet classification as a rearrangement of the thesaurus.30/

30/ Derek Austin, a critical review of R.D.C. Vernon and Valerie Lang, "The London classification of business studies", Journal of Documentation, vol. 27, No. 1, London, March 1971, p. 57.

/6.1 The

6.1 The Leyden thesaurus [31]

121. From 1966, professor van de Merwe began developing the so-called Leyden system which is a documentation system for sociological research methods and techniques.

122. The system, which is still in the development stage and is being prepared for computerized retrieval, has the following main categories:

(a) General methodology
(b) Research methodology. Designing the research
(c) Types of research
(d) Selection of research units
(e) Data collection
(f) Data processing
(g) Statistics. Univariate and bivariate analysis
(h) Scaling. Multivariate dimension analysis
(i) Analysis. Causal analysis
(j) Reliability and validity
(k) Organization of research
(l) Research and researcher in society

Each of these categories is broken down into numerous facets, which permit a great variety of terminological combinations, imparting great flexibility to the system.

31/ This system is dealt with in the following publications: C. van de Merwe, Documentation on Methods and Techniques of Sociological Research: A new classification scheme as a basis for a selected card file (Leyden, August 1966), mimeographed, 38 pages. C. van de Merwe, Documentation for Sociological Research Methods and Techniques: A Progress Report: introducing a facet-grouped thesaurus of descriptors (Rotterdam, 1970), 34 pages. C. van de Merwe, and A.M. de Jong-van der Poel, Documentation for Sociological Research Methods and Techniques; Thesaurus: facet grouping of descriptors and unauthorized terms, Part One (categories A-F) (Rotterdam, 1970); and C. van de Merwe and A.M. de Jong-van der Poel, Documentation for Sociological Research Methods and Techniques: Alphabetical Listing of descriptors and unauthorized terms, Part One (categories A-F), (Rotterdam, 1970).

/6.2 The

6.2 The Thesaurofacet

123. The English Electric "Thesaurofacet" is an interesting new departure. It was originally planned as the fourth edition of the "Faceted Classification for Engineering",[32/] but during its compilation it evolved into a species of retrieval language very different from its predecessor, containing a faceted classification integrated with a thesaurus. The English Electric Company was forced to redraft the thesaurus, which had fallen out of date partly because of advances in the rapidly developing fields of electronics and computer engineering, and partly because the Company's interests had expanded to cover a large number of fields in science and technology.

124. The Thesaurofacet contains approximately 16,000 index terms with individual code symbols and 7,000 entry points, covering synonymous forms of index terms.

125. In order to unite the thesaurus with the facet schedules, the terms appear once in the thesaurus and once in the schedules, and are linked by the notation or class number. At each location different information is given about the term; thus, the two parts of the system are complementary and are incomplete if consulted apart.

126. The third edition of the English Electric Faceted Classification divided the subject-field of engineering into fundamental facets of personality, matter, energy and space, but the operation of the system showed that, in practice, the imposition of an arbitrary and subjective framework on a wide subject area brings many disadvantages.[33/]

127. First, there was no place in the scheme for the traditional breakdown of engineering into, for instance, mechanical, electrical or civil engineering, since the constituent parts of those disciplines were scattered through the primary facets. Secondly, some subject-fields, such as "management" or "mathematics", may have multiple "atribute" and "entity" facets - but they

---

32/ See J. Binns and D. Bagley, A faceted classification for engineering. (English Electric Company Ltd., Whetstone, Leics.), third edition, pages 95-115.

33/ Jean Aitchison, "The Thesaurofacet", in Journal of Documentation, vol. 6, No. 3, pages 187, 188, 195 and 196 (September 1970).

/are fixed

are fixed under one attribute facet,"operations", in the scheme. Thirdly, some concepts could be classified as either one facet or another depending on the point of view.

128. When the main class structure of the Thesaurofacet was being considered it was decided to return to the grouping of subjects under traditional disciplines and to use facet techniques only within those disciplines, retaining the distinction between sciences and technology.

129. The multidisciplinary affinities of concepts are displayed principally through the thesaurus and the cross references, which show that a concept has more than one place in the field of knowledge. Given the complexity of the system, the Thesaurofacet abounds in errors, "many of which came to light after the volume was in the printer's hands and correction impossible without incurring heavy expense... Funds for the project were sufficient only for a small team of compilers /during/ the twenty-month period of compilation".[34/]

6.3 The English classification of business studies

130. The London classification of business studies[35/] - which contains about 2,300 terms - shares some points of resemblance with the Thesaurofacet, although it lacks that essential interdependence between the schedules and the index. Concepts in the field of business studies are organized so that associations between them are displayed on the page through indentation and by a system of expressive and hierarchical notation. Although a preferred citation order is indicated, this is not mandatory, and other formulae could also be used in the building of compound subjects.

131. The classification is organized under three main heads:

(a) Management responsibility in the enterprise, which includes such factors as finance, personnel, office services, production and marketing;

(b) Environmental studies, covering economics, law, communications, etc.

(c) Analytical techniques, including operations research, statistics, automation and workstudy.

34/ Ibid., p. 202.

35/ See R.D.C. Vernon and Valerie Lang, The London Classification of Business Studies, Graduate School of Business Studies, London, 1970.

/6.4 The

6.4 The ISIS thesaurus

132. A brilliant and very recent solution to this problem is the ISIS system developed by the ILO, which considers each set of Boolean operands within parentheses as a facet, i.e. a group of terms - or descriptors - united by "or" in a search request. Each group may be "and"ed with other groups or elements. In this way, the facet concept - which is now endowed with maximum flexibility, since the facets are created at will, in accordance with the user's needs - is united with the Boolean computerized retrieval technology, resulting in a powerful integration.

133. An example of this concept is provided by the following formula:

(A*B*C) ↓ (D*E) ∴ F # G

where the sign * denotes the "or" operation, the sign ↓ the "and" operation and the sign # the excluding "not" operation.

The elements ABC constitute the first facet for the search, the elements DE the second facet, and so on.

It should be borne in mind that the "and" and "not" operations serve to reduce the number of matches, while the number is increased by the "or" operation. Thus, the facets operate to expand the search until the result desired by the user is reached.[36/]

7. International standardization of classification

134. The growing interdependence of scientific and technological knowledge and economic and social phenomena calls for the adoption of standardized terminologies and uniform systems for classification and transfer between the fields of learning and the most important languages.

135. As the Jackson report recognizes, "Fundamental to the development of any information system is effective categorization or classification of the data",[37/] and it is also essential that the classification should be defined and applied uniformly, either in the United Nations system, as indicated in the report, or in any other global system.

36/ ISIS, A General Description of an Approach to Computerized Bibliographic Control, International Labour Office, Geneva, 1971, page 19.

37/ A Study of the Capacity of the United Nations Development System, (United Nations publication, Sales No.: DP/5), vol.II, chapter six, paragraph 44.

/136. However

136. However, "it would not be imperative to have absolute uniformity below the country-sector and major-type activity, particularly if computers are available to translate from one classification scheme to another. Nevertheless, it will be absolutely essential, if different classification schemes continue to prevail, to make certain that they are compatible down to the level suggested and that translations of detailed data below that can readily be made. Further, it will be desirable to work toward consistent formatting of the uniformly classified data in order to provide a firm base for evolution toward a completely integrated, fully automated system to serve all components of the UN system." 38/

137. The Henderson report is in total agreement with the Jackson report in its repeated insistence on the need to establish compatible and common or collective information systems. "In our report continuous emphasis is placed on the urgent need for interorganizational co-operation and co-ordination in data processing information systems." (Author's underlining.) 39/

138. The report also considers that compatibility of equipment is not a serious problem, as long as high-level computer languages are used, while incompatibility of systems has a much more serious impact on effective and efficient use of total resources, since lack of uniformity in format and content of data files hampers the easy exchange or consolidation of information. 40/

139. Differences in the scope of the information stored have led to vast differences in computer programmes; consequently, incompatibility between systems has arisen and a disproportionately large investment has ensued in the development of document information storage and retrieval systems. If efforts had been pooled, "it is reasonable to expect that at the very least, a core system could have been developed to serve all, that compatibility of systems between organizations would have been assured, and that for the same investment or less, a more advanced system of document information storage and retrieval would be available today". 41/

38/ Ibid. chapter six, paragraph 46.

39/ Electronic data processing, op. cit., page I-8

40/ Ibid., page II-6

41/ Electronic data processing, op. cit., pages V-3 and V-4

/140. Freiherr

140. Freiherr von Ledebur states that "the preconditions which Henderson lists for an information system serving a variety of user groups ... are (a) unification of title entries, (b) unification of category contents, (c) standardization of terms, and (d) unification of evaluation methods".[42/]

8. The OECD Macrothesaurus

141. The Organization for Economic Co-operation and Development (OECD) has been working since 1966 [43/] on the preparation of a common list of descriptors for economic and social development, with the co-operation of the International Labour Office, the International Committee for Social Sciences Documentation, the Food and Agriculture Organization of the United Nations, the German Foundation for the Developing Countries and other organizations. The work has thus been a co-operative effort by a group of international organizations.

142. Fourteen international or national organizations have taken part in the tests of the descriptors, and initially some 2,500 documents were analysed in different centres. The computer analyses made it possible to determine the frequency of use of the descriptors, the proportion of new terms to old ones, the kinds of associations made between uniterms to form compound terms, and the pertinence of the answers in the light of the number of documents which had been recorded and the search strategy employed. The results of all this research were incorporated in the new edition of the Aligned List of Descriptors in 1969.[44/]

143. The 1969 edition consists of five volumes, the first of which contains the structured list, in which descriptors are arranged, according to their significance, in primary groups or semantic fields. This volume has to be used to obtain an over-all view of the descriptors available in a particular field. If it is not used, there is a risk that very general descriptors will be used instead of specific descriptors that are more appropriate to the topic dealt with.

42/ Freiherr von Ledebur, Build-up and Organization of a National Information Network; paper presented at the third FID/CLA Congress, Lima, September 1971, mimeographed, page 10.

43/ OECD, Aligned Descriptor List for Economic and Social Development (Paris, 1966).

44/ OECD, Aligned List of Descriptors (Paris, 1969), 5 volumes.

/144. The

144. The structured list is arranged in the following ten major semantic fields:

I Documentation - Research - Methodology
II Chemistry - Physics - Earth Sciences
III Biology - Food - Pathology
IV Agriculture
V Industry
VI Labour
VII Economics
VIII Culture - Society
IX Administration - Politics
X History

145. The alphabetical lists, which are different for each of the three languages (English, French and German) refer to the page and line of the structured list which contains the descriptors. If the analyst consults these lists only and not the structured list, he runs the risk of limiting his vocabulary and using ambiguous descriptors,with the result that information is lost and the computer will come up with the wrong documents. Since the Aligned List is multilingual it is possible to process in one language documents written in several languages, thus ensuring that the synopsis will be consistent with the original text as regards terminology.

146. Lastly, the volume of annexes contains the list of descriptors for countries and institutions and should be used as supplementary information.

147. The synopses, in which the descriptors are shown between slahes, are prepared in ordinary language, using connecting words that are not retrievable, which makes for a clear and concise text. This method has the following advantages with regard to indexing by keywords:

(a) The synopsis produces a legible text;
(b) The keywords are clarified by the fact that they are related in a sentence;
(c) The total is different in nature from the sum of the descriptors that would result from indexing;

/(d) Experience

(d) Experience shows that synopses generate more descriptors than actual indexing;

(e) Synopses highlight the significance, while indexing gives priority to words.

148. At the Meeting of the Group of Experts Aligned List of Descriptors, which was held in Geneva in December 1970,[45/] it was decided to revise the 1969 edition drastically on the basis of a computer analysis of the data collected by OECD, in order to determine the content of the common thesaurus or macrothesaurus in accordance with the following rules:

(a) To keep in the macrothesaurus all descriptors proposed by at least three organizations and to add those proposed by four or more;

(b) To establish a maximum of 35 characters for each descriptor, eliminate abbreviations (except CONF and PUB), use the singular unless the plural has a different meaning, and avoid hyphens between words;

(c) To establish a single numerical country code (United Nations standardized code) with generic descriptors, such as, Africa, South America, etc.;

(d) To establish a revised list of descriptors for economic and social development institutions;

(e) To continue using the classification in broad semantic fields; to reorganize it, identifying each of the groups by means of a numerical code; to relate this classification to that established by the United Nations Consultative Committee on Administrative Questions, with a view to associating documental and operational activities in the development field; and to relate the specialized lists of descriptors of the participating organizations at a given specific level.

45/ This meeting was attended by representatives of the following organizations: UNIDO, UNCTAD, Conférence Européenne des Ministres des Transportes, ECLA, FAO, Instituto Superior de Ciencias Sociais e Politica Ultramarina (Portugal), Agency for International Development, Deutsche Stiftung für Entwicklungsländer, the Dag Hammarskjöld Library at United Nations Headquarters, UNESCO, ILO, World Council of Churches, UNDP and International Development Research Centre of Ottawa. Mr. Viet, Deputy Secretary-General of the International Committee for the Documentation of Social Sciences, acted as Adviser-Rapporteur.

/149. Stress

149. Stress was laid on the co-operative character of terminology control and the need to establish an open system in which new descriptors could be included by those participating in the system.

150. In the light of the experts' recommendations, the President of the OECD Development Centre decided that a new edition of the Aligned List of Descriptors should be prepared, incorporating the experience obtained and the considerable number of vocabularies elaborated by various organizations on the basis of the Aligned List of Descriptors, and that radical changes should be introduced in the content and orientation of the list to reduce its vocabulary and organize it along stricter lines in order to obtain the elements of a language which will cover all the essential branches of development that are interrelated within the framework of a genuine thesaurus.[46/]

151. Figure I shows the possible relationships between the common thesaurus or macrothesaurus and the specialized thesauri of organizations of the United Nations system and external organizations.[47/]

152. The area of application of the Aligned List of Descriptors has gradually spread, generally complementing the specialized lists of descriptors and thesauri of the various organizations. The Aligned List of Descriptors is being used or tried out by a wide range of institutions such as the ILO, FAO, UNESCO, UNCTAD, UNIDO, IDB, AID and IICA-CIDIA of OAS, and has recently been accepted for use in the United Nations PRIS sub-system. The manner in which it has spread has revealed its advantages and disadvantages and provided a broad experimental base for preparing a macrothesaurus.

153. The OECD macrothesaurus seems to be the best means that could be adopted at present to arrive at the standard classification system referred to in the Jackson report and to "build on and utilize fully the information system infrastructure now in place, or being developed, throughout the United Nations

---

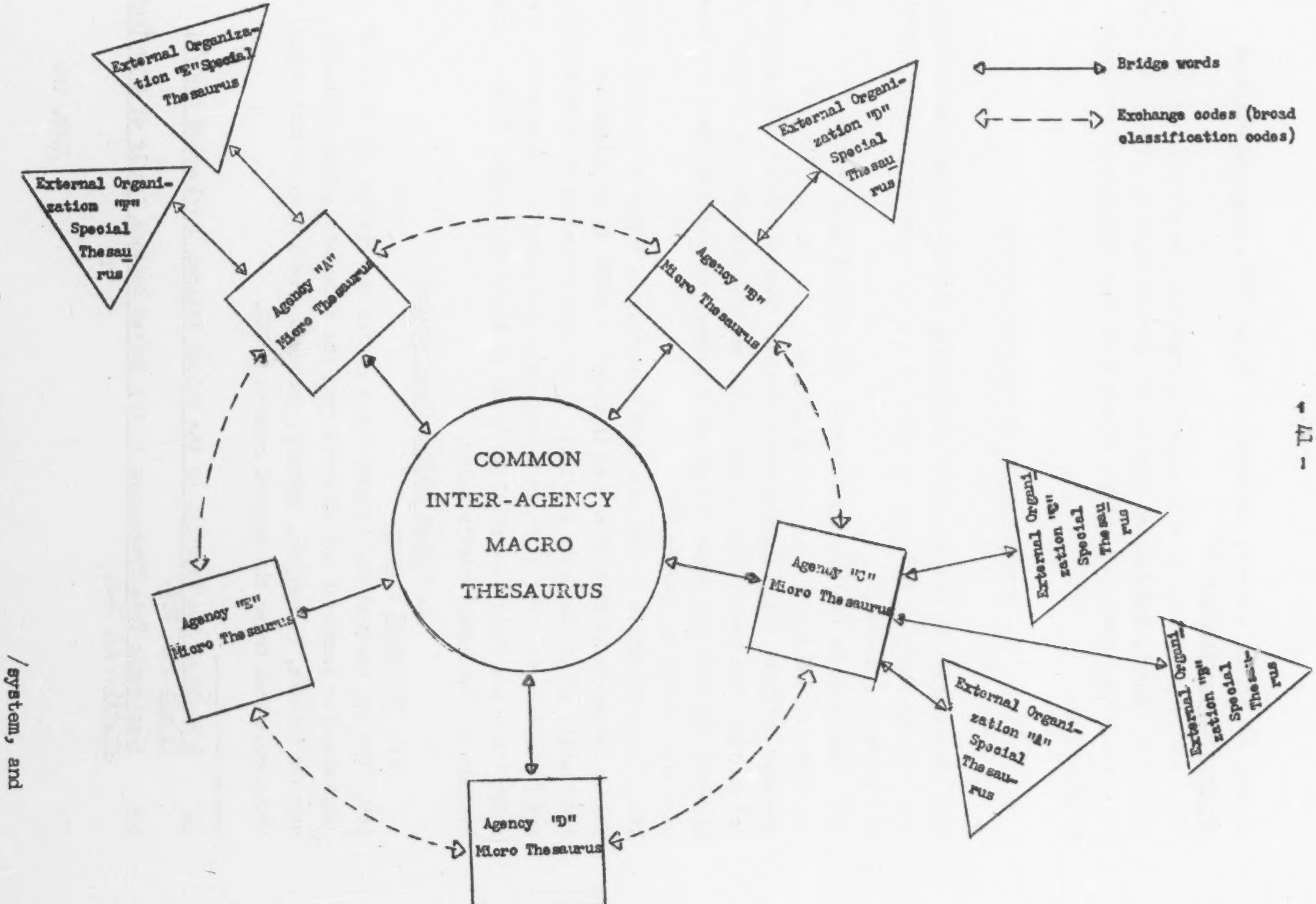
46/ See OECD, Information and Transfers of Experience Division, Development Centre, Meeting of the Group of Experts Aligned List of Descriptors, Main conclusions, 7-8 December 1970 (CH/AD(70)4 and CD/AD(70)1).

47/ Conceptual design of terminology network distributed at the Meeting of Experts by its author, Mr. Giuseppe Martini, Chief, Documentation Division, Dag Hammarskjöld Library, United Nations Headquarters.

/Figure I

Graphic 1

CONCEPTUAL DESIGN OF TERMINOLOGY NETWORK


Bridge words
Exchange codes (broad classification codes)
COMMON INTER-AGENCY MACRO THESAURUS
Agency "A" Micro Thesaurus
Agency "B" Micro Thesaurus
Agency "C" Micro Thesaurus
Agency "D" Micro Thesaurus
Agency "E" Micro Thesaurus
External Organization "E" Special Thesaurus
External Organization "n" Special Thesaurus
External Organization "D" Special Thesaurus
External Organization "C" Special Thesaurus
External Organization "B" Special Thesaurus
External Organization "A" Special Thesaurus

/system, and

system, and in key external organization (e.g., OECD, Deutsche Stiftung für Entwicklungsländer)".[48/]

154. The language area is also expanding rapidly. In addition to the original version in French, English and German, the Spanish version prepared by CLADES has already been completed and the translation into Portuguese and Arabic is under way.

F. UNITED NATIONS INFORMATION SYSTEMS

155. In addition to Headquarters, practically all the specialized agencies have, or are planning to have, information systems in which lists of descriptors or key words are used.

156. According to the Henderson report, the systems used by the Dag Hammarskjöld Library at Headquarters, FAO, the ILO and IAEA are the most advanced and have the following similarities: they all store bibliographic information, use descriptors or key words, produce indexes, are largely incompatible with the other systems with respect to computer input and have been very costly to develop.[49/]

157. It is essential to obtain a good knowledge of these information systems and an interrelationship with them in order to build up an integrated information system in Latin America. It is also necessary to have Spanish and Portuguese versions of their vocabularies and lists of descriptors, for which there are still no equivalent terms in these languages. Some of these systems are described briefly below.

1. United Nations Headquarters

1.1 The UNDEX system

158. The Dag Hammarskjöld Library has a system for storing the texts of comprehensive summaries and extracts from its documents, which produces various indexes, by subject, country, document, resolution of main organ, etc., published under the general name of UNDEX.

48/ A Study of the Capacity of the United Nations Development System, op.cit., p. 238.

49/ Electronic Data Processing in the United Nations Family of Organizations, op.cit., pp. V-2.

/159. The

159. The information is mechanically selected, extracted and compiled in Spanish, French, English and Russian on the basis of notes prepared only in English, which are stored in the computer together with the lists of terms in four languages. The terms consist of groups of words in capitals - which describe the main subject of the document - and, where necessary, words in lower-case letters which relate to details of the principal subject. The complete texts are kept in both their original forms and in microfiche.50/

1.2 Terminology lists

160. United Nations Headquarters has also started to publish computer terminology lists for internal use.

1.3 The Project Reports Information Subsystem (PRIS)

161. An experimental subsystem for Special Fund and other technical assistance projects called Project Reports Information Subsystem (PRIS) has been initiated recently. It will consist of computer lists (bibliographic, of extracts, country indexes, subject indexes, etc.) linked to a system for reproducing projects in microfiche.

162. This system, which uses the techniques developed by the United Nations Library, contains summaries and principal and additional supplementary descriptors, in addition to bibliographic and control data. The principal descriptors relate to the subjects dealt with in the report, and the additional descriptors relate to its content.

2. The ILO Integrated Scientific Information System (ISIS)

163. The ILO Integrated Scientific Information System is the most developed of all those in use by United Nations agencies. Established in 1964 as a documentation system for mechanically processing document analyses, it later became a computer-operated system, based on discs, which can be consulted from remote terminals of a cathode ray screen connected to the computer. The addition of new records or the correction of existing records is also effected through the terminal station. The synoptic descriptors of each document are stored in the same order as they appear in it, between slashes, so that they can be identified by the electronic computers.

50/ UNDEX, United Nations Documents Index, Series A, Subject Index, and Series B, Country Index, from 1970 onwards (United Nations, New York).

/164. The

164. The format used for the storage of records is a modification of the Marc II (Machine-Readable Cataloguing Project, of the United States Congress Library); the version adopted as being the most flexible is the Cox and Davies "Newcastle standard".[51/]

165. ISIS consists of several modules, including document analysis and bibliographic control. The system consist of over 40,000 bibliographic records in the economic and social development field, each record containing a summary prepared on the basis of selected descriptors in a controlled vocabulary or thesaurus. The special type of co-ordinated indexing used in this system enables the user to obtain the records in response to his question through a group of well-defined subject terms. The material includes a collection of 1,000 development projects from countries all over the world.

166. One of the advantages of this system is that some of its modules, programmes and operations are shared by the systems of bibliographic control, regular publications, loans and evaluation.[52/]

3. FAO Documentation system

167. The FAO system links key words and descriptors to form indexing sentences which describe documents. FAO plans to establish a Documentation and Information Retrieval System (DIRS), a computerized system designed to expand its current system of documentation, which covers reports and other documents published by the organization, apart from providing bibliographic lists, lists of authors and KWIC indexes. FAO has a thesaurus of more than 8,000 descriptors and key words which, together with other words, are used to form synopses of the contents of documents. The system in use consists of eleven computer programmes in Cobol and contains information on more than 24,000 FAO documents.[53/]

51/ See N.S.M. Cox and R.S. Davies, "On the communication of machine processable bibliographic records", Programme, vol. 4, N° 3, July 1970.

52/ For a description of the Integrated Scientific Information System, see G.K. Thompson et al, ISIS, A Short Guide for Librarians and Documentalists (LD/Notes/50), International Labour Office, Geneva, 1970; ISIS, A General Description of an Approach to Computerised Bibliographic Control, op.cit., International Labour Office, Geneva, 1971.

53/ Electronic Data Processing, op.cit., page V-3

/4. UNESCO

4. UNESCO system

168. UNESCO plans to introduce a computerized documentation service for indexing and, perhaps, for retrieval, based on the ILO system. The system will cover about 20,000 Technical Assistance experts' reports that are not yet indexed by subject or by key words, in addition to other reports and official documents. It is estimated that total growth in volume in both types of documentation will be about 10,000 documents a year.

169. UNESCO also plans to maintain multilingual terminology lists in four languages.[54/]

170. The UNISIST system of UNESCO and of the International Council of Scientific Unions is aimed, among other things, at establishing a world register for scientific and technical journal titles and abbreviations and at conducting a pilot study to determine the best method of document selection and thesaurus display for scientific and technical thesauri.[55/]

5. World Health Organization

171. The WHO has several information systems, such as the Medical Literature Analysis and Retrieval System (MEDLARS), which it shares with the National Library of Medicine (United States), the WHO Biomedical Research Information Service (WHOBRIS) and the Programme Information Retrieval System (PIRS).

172. The MEDLARS system is the most comprehensive, and contains about 10,000 descriptors which are used for the retrieval of about two and a half million references. The system is entirely in English and has not been adapted to any other language; this raises problems of compatibility with Spanish, which are accentuated by the fact that many names of diseases are not the same in different parts of the Spanish-speaking world.

6. International Atomic Energy Agency

173. The IAEA has the International Nuclear Information System (INIS), which was initiated in 1970. The INIS involves the storage and retrieval of nuclear bibliographic data produced by Member States signatory to the Nuclear

---

54/ Electronic Data Processing, op.cit., pages A-16 and A-17.

55/ See UNISIST, op.cit., and Electronic Data Processing, op.cit., pages A-18 and A-19.

/Non-prolifération Treaty.

Non-proliferation Treaty. The IAEA uses a system of key words and bibliographic descriptors to identify prepared abstracts. Member States provide data on magnetic tape, punched paper tape or typewritten work sheets. This input is processed by IAEA to produce indexes and lists. Microfiche abstracts are also prepared. The EURATOM thesaurus will be used as the basis for the IAEA thesaurus.[56/]

7. International Telecommunication Union

174. The ITU has several information systems, including the Technical Monitoring System, which summarizes measurement data contained in the documents that it receives at the rate of 1,000 a day from radio monitoring stations throughout the world. The Union's International Register of Telegraph Offices contains the names and location of some 300,000 telegraph offices.[57/]

8. World Meteorological Organization

175. The WMO has developed a glossary of technical terms and definitions in Spanish, French, English and Russian, that is specific to hydrological terms, although expansion of the system is planned to include meteorological terminology The system has been developed as a computer file, so that glossaries can be published in alphabetical sequence and cross-referenced using the Universal Decimal Code. The World Weather Watch System comprises four subsystems which process information from land stations, ships and observing stations.[58/]

9. United Nations Industrial Development Organization . (UNIDO)

176. UNIDO has organized an industrial information service which provides practical assistance to developing countries in the form of free answer to their questions concerning industrial problems. This service uses the resources of an information system set up by national and international industrial development organizations, in addition to its own documentation centre and specialized collections of data.

177. By means of its international collection of industrial directories it can furnish lists of equipment suppliers, and advice regarding prices, quality, and delivery and payment terms, and in the formulation of

---

56/ Electronic Data Processing, op.cit., page C-10.

57/ Electronic Data Processing, op.cit., pages F-9 and F-10.

58/ Electronic Data Processing, op.cit., pages G-13 et seq.

specifications and the evaluation of offers. It also possesses collections of data on industrial opportunities in developing countries which require foreign technical know-how and investment, and keeps a register of consultants specializing in industrial development. UNIDO has a list of descriptors based on the OECD list.[59/]

10. International Bank for Reconstruction and Development (IBRD)

178. The World Bank has no information centre or centralized data bank, but since 1965 it has been setting up a data bank for the compilation of world statistics tables; the data are being processed and analysed by the computer centre of IBRD and of the International Monetary Fund. The Bank has economic and social data on 140 countries, including 27 in Latin America. These data comprise 271 time series on population, labour force, national accounts, investment, government finances, international finance, production, prices and balance of payments.

179. The World Bank has fifty computer programmes which are used for data processing and analysis, projections, etc. The programmes are written in Fortran, Cobol or Algol, and include a series for the establishment of a data bank of macroeconomic indicators by country, indicator and year.

180. The following economic and social categories are used in the data bank: 1. population and area; 2. labour force; 3. employment, income and wages; 4. education; 5. health and demography; 6. urbanization and housing; 7. gross domestic product and national accounts; 8. national income distribution; 9. general sector analysis; 10. manufacturing; 11. agriculture; 12. energy; 13. investment and savings; 14. government finances; 15. money supply and credit volume; 16. prices and cost of living; 17. production and value added; 18. balance of payments; 19. international trade of Latin America; 20. interregional trade; 21. unit volumes and quantum of trade; 22. international finance and capital flows; 23. external debt; 24. exchange rates and conversion factors; 25. gross domestic product deflators.[60/]

59/ See UNIDO Industrial Information Service and UNIDO Newsletter.

60/ Personal research and communications.

/11. Inter-American

11. Inter-American Development Bank project

181. The Inter-American Development Bank (IDB) plans to establish an economic and social data system, as far as possible in line with the United Nations and OECD specifications.

182. To establish this system, it is preparing a master list of economic and social categories which will be subdivided into data series organized on the basis of the above-mentioned classification system. The data series will be computer processed with a view to obtaining analyses and projections from 1950 onwards.

183. IDB is also carrying out a bibliographic project on transport, by computer, the first results of which have already been obtained.[61/]

12. The UNCTAD system

184. UNCTAD and the United Nations Economic Commission for Europe propose to establish a joint information retrieval system. As mentioned elsewhere in this document, the two institutions are studying the application of a standardized product code.

G. THE CLADES SYSTEM OF DESCRIPTORS

185. Any regional information system should be closely related to universal systems, such as those of the United Nations and its specialized agencies, referred to above, without prejudice to their possible adaptation to reflect as exactly as possible the situation, problems and needs of Latin America, especially as regards economic and social development.

61/ Personal communications.

/186. In

186. In order to contribute to this work, CLADES is studying the possibility of adopting a system of economic and social descriptors - for eventual conversion into a thesaurus - which would be an effective instrument for retrieving information in spheres of interest to ECLA, ILPES and other regional agencies and, generally speaking, international organizations, governments and research workers interested in the development of the region.

187. The system is being drawn up in accordance with the following methodological guidelines:

1. Objectives

1.1 Efficiency

188. The system is to be used for the retrieval of the greatest possible quantity of information to satisfy each request; unsuitable information should be discarded as far as possible. In this connexion, the experiences mentioned in the previous section should be borne in mind.

1.2 Simplicity

189. The system of selecting and arranging descriptors should be as simple as possible to allow for the co-operation of as many users as possible and to facilitate the work of documentalists.

1.3 Low cost

190. The system of selecting, organizing and processing descriptors should be of the lowest possible cost, which means that it should be of rapid and easy application, without the need for a large or highly specialized staff, and such manual, mechanical and electronic techniques should be used as provide for maximum economy.

1.4 Compatibility

191. The system should have the greatest possible compatibility with the systems already in operation or planned by the United Nations, and particularly with the Headquarters UNDEX and PRIS systems, the ISIS system of the ILO, the systems of regional organization like the IDB, and others of similar potential.

1.5 Standardization

192. As far as possible, the system should meet internationally accepted norms, such as those of UNESCO, ISO, etc.

/2. Development

2. Development model

2.1 Translation of the Aligned List of Descriptors

193. CLADES has completed the translation of the OECD Aligned List of Descriptors, which already exists in three other languages and provides a preliminary nucleus of internationally compatible descriptors. This vocabulary, particularly in the economic and social sections, is one of the terminological modules of the system.

2.2 User profiles

194. In April 1971, CLADES carried out a study of the ECLA Division in their capacity as users of documentation. The basic aim was to determine the interests of users, to evaluate in general terms the frequency with which they had recourse to various sources of information, to discover what documentation was available to them independently of the Central library and to obtain lists of the main subject categories to be found in the Division libraries, as an expression of their specific interests.

195. A series of basic candidate descriptors is being drawn up from the replies of the Divisions and from the analysis of their work programmes,[62/] and their value for documentation retrieval will be tested in a short while.

196. Figure 2 shows the global distribution of sources of information and the relative frequency of their use. Of course, this macroanalysis is no more than a preliminary approximation which will have to be elaborated in the future.

2.3 United Nations material

197. The basic material from which CLADES is selecting descriptors for ECLA is derived from the very structure of the Commission, its programme of work and publications, all of which will be supplemented by similar material from ILPES and other regional institutions within the United Nations system.

198. From a structural point of view, ECLA is interested in the following conceptual spheres or categories: Public administration, Agriculture, Technical assistance, Social affairs, Economic development, Industrial

---

62/ See ECLA, Draft Programme of Work and Priorities (E/CN.12/875/Rev.1).

/Figure 2

Figure 2

OVER-ALL DISTRIBUTION OF ECLA SOURCES OF INFORMATION BY SUBSTANTIVE UNITS AND RELATIVE FREQUENCY OF USE

Source:	Public Administration				ECLA/FAO Agriculture				Social affairs				Economic development				Industrial development				Statistics				Special Studies				Export promotion				Trade policy				Economic projections				Natural resources and energy			
	F	R	I	N	F	R	I	N	F	R	I	N	F	R	I	N	F	R	I	N	F	R	I	N	F	R	I	N	F	R	I	N	F	R	I	N	F	R	I	N	F	R	I	N
United Nations	x						x		x					x				x			x						x				x				x						x		
United Nations agencies		x			x				x					x				x				x			x				x				x				x						
Governments			x			x			x						x		x				x				x						x		x				x				x			
Universities and research centres			x					x	x						x				x				x				x							x						x			
Professional associations			x					x	x						x				x				x				x							x						x			
Commercial publishers			x			x			x					x					x			x				x					x			x						x			
Form:																																												
Mimeographed	x					x			x					x			x				x						x					x				x				x			
Book		x					x		x					x					x		x						x					x				x					x		
Periodical	x				x				x					x				x			x						x					x				x					x		
Analytical summaries			x			x					x				x					x			x				x							x						x			
Press clippings			x				x					x			x					x				x				x							x						x		
Other			x					x		x				x				x					x					x							x						x		

Key: F = Very frequent
R = Frequent
I = Infrequent
N = Not used

development, Energy, Statistics, Export of manufactures, Trade policy, Natural resources and Transport, and shares with ILPES an interest in Planning and Economic programming, all of the above with special reference to Latin America.

199. As regards its work programme, ECLA has permanent interests and others which fluctuate from year to year, according to the interests of its research workers. For instance, concern for the human environment as a topic of research is very recent. It calls for a basic vocabulary of descriptors, on the one hand, and the updating of that vocabulary to take account of new problems requiring the attention of the Commission, on the other.

200. The publications of ECLA, ILPES and other United Nations agencies are the most important source of descriptors, since the retrieval of those documents is fundamental to the work of ECLA.

2.4 Government material

201. Such material is another basic source of descriptors, although it is predictable that it will not add much to material from the sources mentioned above, since the interests of ECLA, and of the United Nations in general, coincide with the interests of governments.

2.5 Other material

202. Works relating to Latin America in the economic and social field from universities and other institutions and from commercial publishing houses, are another valuable source of material, particularly in those fields where scientific and technological innovations are produced that are worthy of being incorporated in the terminological system.

3. Organization

203. The work will be organized and developed by stages along the following lines:

3.1 List of descriptors

204. Once the process of selecting candidate descriptors on the basis of the above-mentioned materiales is completed, they will be arranged in an alphabetic dictionary and in semantic fields which reflect the users' interests.

/205. The

205. The list of descriptors will be supplemented by country, institution, product and other codes which may be necessary in practice.

3.2 International co-operation

206. Consultations are being held with Latin American institutions and experts that are experienced in this field and are working with descriptors, in order to establish the regional validity of the use of the descriptors collected and to prepare lists of synonyms. IICA-CIDIA has translated the agricultural terms in the Aligned List of Descriptors into Spanish, and co-operation is also expected from other specialists in various fields.

3.3 Disciplinary modules

207. Once the terms have been checked, disciplinary modules, i.e., specialized vocabularies of disciplines that are of relevance for economic and social development in Latin America, will be established and organized in such a way that they can be very flexibly integrated at the level of missions. By this modular system it will be possible to deal not only with the activities under way but also with future missions or projects, being easily adaptable - through the inclusion of modules - to cover new fields, such as the human environment, procedures for dealing with natural disasters, economic exploitation of the sea, etc.

3.4 Preparation of the thesaurus

208. When sufficient experience has been gained in the use of the List of Descriptors and enough cross-references have been established between their components, the time will have come to organize the thesaurus, which will also be built up in a flexible modular form. The experience acquired by that time, not only by CLADES but also by international organizations, particularly United Nations Headquarters, and other specialized institutions, will determine the final structure of the thesaurus and its connexion with other lists of descriptors and existing thesauri.

3.5 Computer programmes

209. CLADES is planning to establish a library of modular computer programmes, which will be at the service of interested governments and institutions.

/210. A

210. A start has already been made on preparing the modules, a first programme being available for the preparation of a dictionary of descriptors and their retrieval, and a second programme for the preparation and retrieval of bibliographic lists, which are operational and are at present being used on an experimental basis. These programmes are very similar to the programmes used in the ILO and OECD systems and PRIS; they have been established in PL-1 language so as to make them as consistent as possible with PRIS, in which this language is used. However, it would not be a difficult matter to establish versions in other languages, such as COBOL and ASSEMBLER, in order to facilitate the work of the institutions using the programmes.

3.6 Codes

211. As regards coding systems, use is made, in principle, of the code of countries members of the United Nations, although it is not considered practical because of the three-digit representation of countries in English alphabetical order, instead of having series for the different continents or geographical groups; for example, 100 for Europe, 200 for Latin America, etc. Secondly, the digits are used as prefixes in that code, which makes it necessary to use five digits in all; this could be avoided if the two additional digits, and others which may be added as suffixes, were used after the country code.

212. Once this problem is discussed, the United Nations may well adopt a new approach to country coding, which would also be adopted by CLADES.

213. ECE is establishing a code for primary products, on the basis of the Brussels Tariff Nomenclature (BTN) and the uses of the Customs Co-operation Council, which will probably have eight digits.

214. Wherever necessary, other codes will be used, as in the case of the internationally standardized currency and vehicle licence plate codes.

215. EEC and UNCTAD plan to convert the alphabetical licence plate codes into numerical codes for use in computers, and also for other uses such as the identification of containers.[63/] CLADES will use those codes and, once they are internationally approved, will promote their use in Latin America.

63/ Personal information furnished by Mr. Gösta E. Ross, UNCTAD interregional adviser on trade documentation, June 1971.